

Tax fraud by firms and optimal auditing

Article

Accepted Version

Hashimzade, N., Huang, Z. and Myles, G. D. (2010) Tax fraud by firms and optimal auditing. *International Review of Law and Economics*, 30 (1). pp. 10-17. ISSN 0144-8188 doi: <https://doi.org/10.1016/j.irle.2009.08.002> Available at <https://centaur.reading.ac.uk/16928/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.irle.2009.08.002>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Tax Fraud by Firms and Optimal Auditing

Nigar Hashimzade*
University of Reading

Zhanyi Huang
University of Exeter

Gareth D. Myles
University of Exeter and Institute for Fiscal Studies

July 7, 2009

Abstract

Tax fraud is an issue of increasing importance in China. One particularly significant fraud involves excessive claims for the rebate of VAT on exported goods. This fraud has two interesting features. First, it requires the collusion of an intermediary to supply the false documentation that supports a rebate application. Second, the punishment schedule is convex – with capital punishment used in major fraud cases. These features ensure that the payoff function of a firm engaging in fraud is strictly concave in the level of fraud. This gives a well-defined optimization without the need to appeal to risk aversion. We show that the existence of fraud does not affect the real output decision of the firm nor the tax policy of the government. Audit resources can be used to detect firms engaged in fraud as well as the intermediaries who supply false documents. Under reasonable assumptions it is shown that resources should be focused on detecting firms and not intermediaries. Finally, if the government must take action on fraud a convex punishment scheme is shown to be optimal.

**Corresponding author:* Nigar Hashimzade, School of Economics, The University of Reading, Whiteknights, Reading RG66UD, UK, n.hashimzade@reading.ac.uk

The authors are grateful to the anonymous referee whose comments and suggestions had greatly helped to improve the paper.

1 Introduction

In many countries there is growing public concern over issues of tax evasion and fraud in economic activities. Evidence indicates that tax evasion is especially severe in developing and transition economies (Schneider and Enste, 2000). The figures reported by Jia (2002) show that marketization in China has been accompanied by the development of a hidden economy. The economic literature has focused largely upon the evasion of income taxation. In contrast, corporate tax fraud is an important issue in China and a particular concern of its tax system. Tax fraud is a form of economic crime and the source of a potentially serious loss of revenue to the Chinese government.

The Chinese tax system relies on a form of voluntary compliance. Under Chinese law, all taxpayers are required to assist the tax authority by reporting honestly their income and profits and paying a level of tax based on their reported incomes over time. The system is actively enforced by the State Administration of Taxation (SAT) and the Chinese courts which can target investigations and impose substantial penalties for evasion and fraud. In the 2005 tax year the SAT raised revenue of 3,086.6 billion Yuan, or 16.93% of GDP. On the other hand, Jia (2002) estimated the revenue lost through evasion and fraud at 444.6 billion Yuan, which was about 35.11% of the total tax revenue of 1,266.6 billion Yuan in 2000.

In this paper we analyze tax fraud in claims for Value Added Tax (VAT) rebates on exported products. This fraud has some especially interesting features from the perspective of tax theory. Producers can claim a VAT rebate on the value of exports, so overstating the value of exports generates an unwarranted rebate. To commit such a fraud requires a set of supporting documents: a false VAT invoice, a customs manifest, and a foreign exchange settlement memo. A producer committing fraud therefore requires the assistance of another party to supply the documentation. This interaction between the parties is the first interesting aspect of the model. For the VAT fraud the government has a very clear punishment scale with potentially extreme penalties, rising to capital punishment for a sufficient level of fraud. This explicit penalty scheme is the second interesting aspect since it is clearly convex, rather than being the linear fine used in most models of tax evasion.

Many theoretical and empirical studies have addressed the economics of tax evasion. The standard approach was introduced by Allingham and Sandmo (1972) in the context of personal income taxation. They assumed that fines were based on the amount of income evaded by a taxpayer and addressed the determinants of individual income tax compliance. This basic model has been extended in a number of directions which are surveyed in Myles (1995) and Slemrod and Yitzhaki (2000). There is rather less literature on tax evasion by firms. Firms can evade taxation either by misreporting sales or profit, or by making false declarations about input use. It is possible that all these methods may be required simultaneously to disguise evasion if the information gathering process of the revenue service is sufficiently thorough. The tax evasion decision for competitive firms has been analyzed by Virmani (1989), Yamada (1990),

and Cremer and Gahvari (1993). Tax evasion by monopolistic firms has been studied by Marrelli (1984) and evasion by oligopolistic firms by Marrelli and Martina (1988).

Various forms of noncompliance with VAT have been reviewed in Keen and Smith (2006): the authors list under-reported sales, failure to register (“ghost” businesses), misclassification of commodities (as tax-exempt or subject to lower rates) as common for VAT and retail sales tax evasion, as well as forms that are distinct for VAT evasion, such as false claims for credit or refunds, credit claimed on purchases that are not creditable, and bogus traders, or firms set up solely to generate invoices used to recover VAT. Keen and Mintz (2004) show how an optimal VAT schedule can be constructed to balance the tax revenues and collection costs (including both administrative and compliance costs). Das-Gupta and Gang (2003) examine the impact of a particular VAT enforcement technique, namely, matching of purchase and sales invoices, on the enforcement and efficiency of VAT. They find that sufficiently intensive audit can induce truthful reporting but may at the same time distort the purchase-sales and the input-output decisions of the firm. Fedeli and Forte (1999) model a VAT evasion decision as an outcome of bargaining between the firms involved in the VAT chain. In particular, their model predicts that for recovering the lost revenues an increase in the probability of audit is preferable to an increase in the penalty for evasion. Aizenman and Jinjark (2008) investigate how different political and structural economic factors affect the collection efficiency of VAT in a cross-sectional sample of countries, but notably not including China. In these papers a traditional, consumption-based form of VAT is assumed. The unusual feature of VAT in China is that it is production-based and capital goods are included in the VAT base (Ahmad *et al.*, 2004). Furthermore, China employs a separate export rebate system, with the rebate rates independent of rates paid on creditable inputs. Whalley and Wang (2008) analyze the welfare implications of these two features of the VAT system in China using a monetary trade model and argue that the switch to the traditional form of VAT in China will not be welfare-improving. They emphasize the importance of export rebates in their analysis, but do not address the issue of evasion. A similar study is undertaken in Lin (2008). Fisman and Wei (2001) present an empirical case study of tax evasion in China, with the focus on China’s imports from Hong Kong and the evasion of tariffs plus VAT. They show that more revenue is lost for products subject to higher tax rates; furthermore, they find strong evidence of evasion by mislabelling, or misclassification of higher-tax products as lower tax products. Our paper contributes to this literature by presenting a theoretical analysis of fraud involving a VAT rebate for export sales.

The analysis of tax evasion by firms faces a difficulty when combining profit maximization with a linear punishment schedule: expected profit is linear in the amount of evasion, so the decision to evade is all or nothing. Marrelli (1984) and Marrelli and Martina (1988) resolve this problem by assuming the owner of the firm is risk averse. A similar approach is undertaken in DeMello (2008). This is acceptable in the context of an owner-manager but does not fit convincingly with wider ownership of the firm’s equity. Stöwhase and Traxler (2005) assume

instead that there is a cost of evading which is convex in the amount of evasion but do not identify the source of this cost. In contrast, the model we study is necessarily nonlinear for reasons entirely explicable for the form of tax fraud we analyze. First, the announced fraud penalty structure is explicitly convex. We detail this in a brief summary of the Chinese VAT system in Section 2. Second, the act of fraud we consider involves the participation of an additional party who must supply false documentation to support the fraud. Section 3 shows how the cost of securing false documentation provides an explanation of the evasion cost used by Stöwhase and Traxler (2005). Section 4 demonstrates that this structure of costs and punishments implies the value of fraud undertaken by the firm is independent of the tax rate. This produces a separation of the government choice problem studied in Section 5 into two distinct components. The first component determines the optimal tax rate, and this rate is unaffected by the existence of fraud. Fraud does not, therefore, imply that the tax rate is distorted from its efficient value. The auditing and punishment policy is a consequence of the second component. If the marginal cost of auditing firms and document suppliers is the same then resources should be focussed upon auditing firms, not document suppliers. It is shown in Section 6 that the same conclusions hold under an alternative specification of the punishment structure. Section 7 demonstrates that if there is a limit to the level of punishment that can be exacted for a relatively small amount of fraud then convexity of punishment emerges endogenously as the optimal structure of punishments. Section 8 provides the conclusions and some technical details are presented in the Appendix.

2 Tax Rebates

China, like many other countries, encourages exports and provides tax exemptions and other incentives to industries that produce predominantly for the international market. Since 1 January 1994 companies have been able to claim a tax rebate upon exports¹. The level of refundable VAT is equal to the taxable amount times the unit refund rate. The rate is either 17%, 15%, 13%, 6%, or 5%, depending on whether the products are industrial, electrical, agricultural, or other goods. In 2004 about 219.59 billion Yuan in VAT refunds were claimed (SAT, 2005). More recently, from 1 December 2008, the tax rebates were increased for 3,770 items of labor-intensive, mechanical, and electrical products, or 27.9 percent of the country's total exports. A further increase in the rebate rate took effect from 1 January 2009. Chao *et al.* (2008) have analyzed the role of the tax rebate in promoting Chinese exports.

¹To claim the rebate the enterprise must be qualified as falling into one of the following eight categories: Manufacturing enterprises with export rights, Foreign trading enterprises with export rights, Industrial and trading enterprises with export rights, Sino-foreign joint ventures with export rights and joint ventures chain enterprises with export rights, Foreign invested enterprises, Enterprises consigning their export goods to other export agents with export rights, Designated tax rebate enterprises and Specially authorized tax rebate enterprises.

To claim a rebate a transaction must be recorded in account books as an export sale. Three basic documents are required to support a claim:

1. VAT invoice: a primary original certificate is required. The goods involved must be subject to VAT.
2. Customs manifest obtained from a customs office. Goods must have passed through customs and have left Chinese territory.
3. Foreign exchange settlement memo: a full set of documents is submitted at a local bank. China practices a foreign exchange settlement system as well as a system of verification and cancellation of foreign exchange receipts from exports.

A falsified claim for a VAT refund occurs through the submission of false documents to the SAT. According to Jia (2002), the purchase of false documentation only costs 2%-4% of the VAT invoice from suppliers, and conspiracies to obtain payment of false claims amount to 17% of the VAT refund from the SAT. A first report by the official media on export rebate related investigations showed that the Chinese government may have lost 30 billion Yuan in tax revenue due to fraudulent practices (South China Morning Post, 1 July 2001). According to the newspaper, some of the practices that businesses used to cheat include “to set up overseas holding firms to issue fake export invoices and shipping documents; to claim rebates with the help of officials despite selling ‘exports’ within China. In addition, exporters have easy access to fake tax invoices. These practices have extensively abused the VAT tax system since its introduction in 1994, and local governments have shown little support of the VAT system or even were involved in the fraudulent practices”.

The punishments for tax fraud are described in Articles 205 and 206 of the Criminal Law of the P. R. of China. Article 205 states that “whoever falsely makes out special invoices for value added tax or any other invoices to defraud a tax refund for exports or to offset tax money shall be sentenced to fixed-term imprisonment of not more than three years or criminal detention and shall also be fined not less than 20,000 yuan but not more than 200,000 yuan; if the amount of money involved is relatively large or if there are other serious circumstances, he shall be sentenced to fixed-term imprisonment of not less than three years but not more than 10 years and shall also be fined not less than 50,000 yuan but not more than 500,000 yuan; if the amount of money involved is huge, or if there are other especially serious circumstances, he shall be sentenced to fixed-term imprisonment of not less than 10 years or life imprisonment and shall also be fined not less than 50,000 yuan but not more than 500,000 yuan or be sentenced to confiscation of property. Whoever commits the act mentioned in the preceding paragraph to defraud tax money of the State shall, if the amount involved is especially huge, and the circumstances are especially serious, thus causing especially heavy losses to the interests of the State, shall be sentenced to life imprisonment or death and also to confiscation of property.” Article 206

records the same punishment structure for “whoever forges or sells forged special invoices for value added tax.”

The seriousness of tax fraud is reflected in the sentences imposed by the courts. Of particular note is a report by the SAT in November 2001 that 19 people had been sentenced to death and 30 given life sentence. In the two cities of Chaoyang and Puning in Guangdong Province 150 people were convicted for using false VAT invoices to secure 4.2 billion Yuan of illegal gains.

3 Fraud and Documentation

The key feature of the procedure for obtaining a tax refund described in Section 2 is that supporting documentation is required. To make a false claim, and to engage in an act of fraud, a firm must obtain a complete set of false documentation including a false tax return. We assume the firm must buy the fraudulent documentation from an “intermediary” who becomes an accomplice in the act of evasion. This purchase of documents provides an additional element of economic interest to the study of fraud in this context since it requires collaboration between parties. The tax law also recognizes the need for false documentation and prescribes punishment for firms that are caught engaging in fraud and for intermediaries caught assisting in the act.

We model the process of document supply by assuming that the firm and the intermediary negotiate over the price of false documents and, with some probabilities, the firm and/or the document supplier might be caught cheating and punished. The price of the false documentation is determined endogenously by the extent of fraud, the probabilities of detection, and the punishments. The punishments for the firm and the intermediary (the fine for tax fraud) follow the practice of the tax code so are convex in fraud, meaning that the greater is the level of fraud as measured by the overstatement of export, the higher is the rate of punishment.

Consider a firm that produces a single good, Y , with the use of a numeraire produced input, K , and labor, L . Assume initially that the firm acts honestly. The tax rebate is based on the value of exports at the f.o.b. price (SAT 2003, 2005) which we assume is equal to the domestic price. In this case the tax liability and the rebate for exporting are given by

$$T_H = t[p_h Y - K], \quad R_H = r\mu p_h Y, \quad (1)$$

where t is the VAT rate, p_h is the domestic price of the output, μ is the fraction of output that is exported, and $0 \leq r \leq t$ is the rate of rebate. As Whalley and Wang (2007) point out, the rebate under this scheme differs from a credit invoice VAT in that it applies to the full value of exports and not just the produced input component. Denoting the market price abroad by p_e the profit made by the firm when honest is

$$\pi_H = p_h Y - K - wL + [p_e - p_h - g(\mu Y)] \mu Y - T_H + R_H, \quad (2)$$

where $g(\mu Y)$ is the transaction cost per unit of export (this can include transportation costs, *etc.*), and w is the per unit cost of labour. Units of the good sold at home and exported are identical.

The firm can commit fraud by overstating the fraction of output exported. This generates an excessive refund of tax but opens the possibility that the firm is caught and punished. The firm chooses the level of excessive claim taking into account the price that has to be paid for false documentation and the fine that is paid if caught. When the firm overstates its export sales it simultaneously has to understate its domestic sales, since the total output by the firm can be verified.

If the fraction of export sales in total sales is falsely reported as $\mu + m$ ($m > 0$), the claimed refund with fraud is

$$R_F = rp_h [\mu + m] Y = R_H + rp_h m Y. \quad (3)$$

Since tax liability is determined by the value of total output at home prices this remains the same at $T_F = T_H$ (taking Y as given for the present). The total value of fraud is thus

$$E = rp_h m Y = rX, \quad (4)$$

where $X \equiv p_h m Y$ is defined as the extent of fraud.

The firm is caught making an excessive claim for refund with probability q_F . When caught it has to pay back the excessive refund and is also subject to a fine of Φ_F . Alternative assumptions can be made about the determinants of the fine (see Yitzhaki, 1974, for a discussion of this issue in the context of personal income taxation). In what follows we focus on the case where the fine is determined by the extent of fraud, so $\Phi_F = \Phi_F(X)$. The consequences of an alternative assumption – the fine being determined by the level of fraud – are briefly summarized in Section.6.

When the extent of fraud is X the expected profit of the firm is

$$E[\pi_F] = \pi_H + [1 - q_F] rX - q_F \Phi_F(X) - V, \quad (5)$$

where V is the cost of the false documentation required to support the fraudulent claim. The expected gain to the firm from engaging in the fraud is given by

$$E[W_F] = E[\pi_F] - \pi_H = [1 - q_F] rX - q_F \Phi_F(X) - V. \quad (6)$$

What we now wish to do is to establish the price for documentation as a function of the extent of fraud, the probability of detection, and the punishment structure. Given this price it becomes possible to analyze the decision problem of the firm. We determine the price of false documentation by assuming that it is determined by a Nash bargaining process between the firm and the intermediary.

The intermediary receives a payment of V from the firm but also faces a chance of being caught and punished. The intermediary's activity in producing false documentation is detected with probability q_I in which case it has to pay a penalty Φ_I . The fine is determined by the extent of fraud, so $\Phi_I = \Phi_I(X)$

(Section 6 reports the outcome when the fine is determined by the amount of tax evaded). The expected gain to the intermediary is

$$E[W_I] = V - q_I \Phi_I(X). \quad (7)$$

Following Crocker and Slemrod (2005) we assume identical functional forms for the penalty functions, so that from this point onward $\Phi_F = \Phi_I = \Phi$. This assumption is consistent with the Chinese criminal law.

Depending on the nature of the intermediary, alternative assumptions can be made about the relationship between q_F and q_I . We consider two possible situations: the intermediary as a corrupt official, and the intermediary belonging to the underground economy. In the first case the government can monitor the intermediary (say, audit customs offices) independently from the producer. It is plausible to assume that once the intermediary is detected there is a positive probability that the firm will be also audited and caught. The same applies to the firm. In the second case the government can only catch the intermediary if the producer is caught cheating. We capture these two possibilities by modelling the probabilities as

$$q_F = \min\{q + \alpha p, 1\}, \quad q_I = \min\{p + \alpha q, 1\}, \quad 0 \leq \alpha \leq 1, \quad (8)$$

where q and p are the probabilities of being independently audited and caught for the firm and for the intermediary, respectively. Thus, the first scenario (corrupt official) corresponds to $p > 0$. The second scenario (underground economy) corresponds to $p = 0$. Parameter α represents the likelihood that catching one party (firm or intermediary) leads to catching the other.

The bargaining parties take the probabilities of being detected as given and solve

$$\max_{\{V\}} N = (E[W_F])^\theta (E[W_I])^{1-\theta} \quad \text{s.t.} \quad E[W_F] > 0, E[W_I] > 0, \quad (9)$$

where $\theta \in [0, 1]$ is the relative bargaining power of the firm. The necessary condition for the choice of document price is

$$\frac{\partial N}{\partial V} = -\frac{\theta N}{E[W_F]} + \frac{[1-\theta] N}{E[W_I]} = 0. \quad (10)$$

Assuming $q_F < 1$ and $q_I < 1$, (8) can be used to write the solution to (10) as

$$V^* = (1-\theta)[1-q-\alpha p]rX + (\theta[p+\alpha q] - (1-\theta)[q+\alpha p])\Phi. \quad (11)$$

It can be seen from (11) that an increase in fraud (X) implies a higher cost of documentation. The effect of the other parameters depends upon the balance of power in bargaining.

4 Production and Evasion Decisions

This section considers the optimization decision of the firm. The firm has to choose the quantity of the two inputs, the division of output between home

and export, and the degree of false reporting. The optimization takes into account the probability of detection and the dependence of the price of false documentation on the extent of fraud.

The expected profit of the firm is given by

$$E[\pi_F] = \pi_H - V + (1 - q_F)rX - q_F\Phi(X). \quad (12)$$

Substituting for V using (11) profit becomes

$$E[\pi_F] = \pi_H + \theta[(1 - q_F)rX - (q_I + q_F)\Phi(X)]. \quad (13)$$

It is best to begin by considering the optimum quantity of overstatement, m . The first-order condition for the choice of m is

$$\frac{\partial E[\pi_F]}{\partial m} = [(1 - q_F)r - (q_I + q_F)\Phi'(X)]\theta p_h Y = 0, \quad (14)$$

Since $\theta p_h Y > 0$ the sufficient condition for overstatement to take place is that

$$(1 - q_F)r - (q_I + q_F)\Phi'(0) > 0, \quad (15)$$

which requires that the probabilities of fraud being detected and the marginal rate of increase of the fine at zero fraud must not be too great. We assume that this condition is satisfied. The necessary condition can be derived from (14) as

$$(1 - q_F)r - (q_I + q_F)\Phi'(X) = 0. \quad (16)$$

From (16) we can develop an expression for the optimal value of X by writing

$$\Phi'(X) = \frac{1 - q_F}{q_I + q_F}r = \frac{1 - q - \alpha p}{(1 + \alpha)(p + q)}r. \quad (17)$$

This can be solved to derive the solution $X = X(p, q, \alpha, r)$.

The effect of changes in the parameters upon the extent of fraud can be obtained from analyzing the comparative statics of the optimization (see Appendix for the details). The central result is that an increase in the rate of rebate results in increased fraud. A policy designed to encourage the firms to export therefore also provides an encouragement for increased fraud. An increase in either q , p or α reduces the amount of fraud since these raise the probability of the firm being caught.

The analysis of overstatement can now be employed to analyze input use and the division of output between home and export. The output of the firm is determined by the inputs of capital and labor through the production function

$$Y = Y(K, L). \quad (18)$$

For $z = K, L$, and μ the necessary condition for profit maximization can be written

$$\frac{\partial E[\pi_F]}{\partial z} = \frac{\partial \pi_H}{\partial z} + \theta[(1 - q_F)r - (q_I + q_F)\Phi'(X)]\frac{\partial X}{\partial z}. \quad (19)$$

The important property of these necessary conditions can be observed immediately: neither input choice nor the export decision is affected by the possibility of fraud. This follows from the fact that the optimal choice of fraud, (16), ensures the second term on the right-hand side of (19) is zero, so the choice of the optimal quantities is determined by $\partial\pi_H/\partial z$ alone. Furthermore, the necessary condition for optimal overstatement does not depend on K , L , and μ , so can be solved independently of those variables. The decision problem of the firm therefore partitions into a real production decision and a separate fraud decision. Engaging in fraud does not distort the production decision. This result is similar to the one obtained in cross-country profit shifting, as, for example, in Stöwhase (2005).²

Having demonstrated the partition of the choice problem we now look in detail at the real production decision. Differentiation of π_H with respect to μ gives the necessary condition

$$p_e - (1 - r)p_h = g(\mu Y) + \mu Y g'(\mu Y) \quad (20)$$

With positive, non-decreasing per unit export transaction cost the firm's exports are positive when the price abroad is sufficiently high, $p_e > (1 - r)p_h$.

The first-order conditions with respect to K and L are

$$-(1 - t) + Y_K [(1 - t)p_h + \mu\varsigma] = 0, \quad (21)$$

$$-w + Y_L [(1 - t)p_h + \mu\varsigma] = 0, \quad (22)$$

where $\varsigma \equiv p_e - (1 - r)p_h - g(\mu Y) - \mu Y g'(\mu Y)$. Using (20), (21) and (22) imply that the ratio of marginal products is equated to the ratio of factor costs adjusted for the tax

$$\frac{Y_K}{Y_L} = \frac{1 - t}{w}. \quad (23)$$

This is the same necessary condition that would occur without fraud so the opportunity for fraud is not distorting the input choices of the firm relative to the choices made by an honest firm. This implies the output level is also not affected by fraud. Hence, this form of fraud has no consequence for the real production decisions of the firm and causes only a false statement of the activities of the firm.

5 Taxation and Auditing

The literature on tax evasion has discussed alternative representations of the optimization problem of the government. The early literature (such as Kolm 1973) permitted the government to simultaneously optimize all choice variables and, in the context of personal income tax evasion, generated extreme conclusions.³ This motivated a focus on the separation between the central government which

²The authors are grateful to the anonymous referee for this point.

³Summarized by Kolm as “hanging tax evaders with probability zero”.

chooses the tax rate, an independent revenue service that controls auditing policy, and a judiciary that sets the level of punishment (Reinganum and Wilde (1986)). We investigate these issues in this section and in Section 7. In this section we show that there is a natural separation of the optimization in the model of fraud – in particular, the optimal tax rate is independent of the audit strategy and level of punishments (and *vice versa*) – and provide a general characterization of the optimal audit strategy. Section 7 focuses upon choice of the optimal structure of punishments.

Consider the choice of the tax rate taking the effort put into monitoring and the fines levied on the firm and the document supplier as given. Let $C(q, p, \alpha)$ be the cost of monitoring when a probability of detection of the firm $q_F(q, p, \alpha)$ and of detection of the document supplier $q_I(q, p, \alpha)$ are achieved. We assume that the government seeks to maximize the sum of tax and fine revenues net of the monitoring cost.

Under this assumption about the objective of the government, the optimal values of t and r solve

$$\max_{\{t, r\}} E[TR_F] = TR_H - (1 - q_F)rX + (q_I + q_F)\Phi(X) - C(q, p, \alpha), \quad (24)$$

where TR_H is the tax revenue less rebate collected when the firm reports honestly

$$TR_H = T_H - R_H. \quad (25)$$

The necessary conditions for tax and rebate rates obtained from (24) are

$$\frac{\partial E[TR_F]}{\partial t} = \frac{\partial TR_H}{\partial t}, \quad (26)$$

$$\frac{\partial E[TR_F]}{\partial r} = \frac{\partial TR_H}{\partial r} - (1 - q_F)X. \quad (27)$$

The first condition implies that the optimal tax rate is the same in the presence of fraud as it is in the absence of fraud. Therefore, the existence of rebate fraud does not distort the tax policy of the government. TR_H is decreasing in r , so that an increase in the rebate rate always reduces expected tax revenue. This is not surprising since the rebate is costly and, as implied by (14), does not affect output so a higher rebate always reduces revenue. The rate of rebate will therefore be set at a level determined by policy considerations that capture more general objectives outside of the issues modeled here.

The value of the tax rate, t , and the rebate rate, r , are now taken as chosen, and we consider the second part of the optimization. This allows us to assess whether auditing resources should be focused upon detecting firms engaged in fraud or intermediaries supplying false documentation (when the latter can be audited separately from the firms). Consider, first, the scenario of the intermediary belonging to the hidden economy so that $p = 0$. In this case only α and q can be chosen. Using (8) in (24) and differentiating with respect to q and α

gives the following expressions for the derivatives of revenue:

$$\frac{\partial E[TR_F]}{\partial q} = rX + (1 + \alpha) \Phi(X) - C_q(q, 0, \alpha), \quad (28)$$

$$\frac{\partial E[TR_F]}{\partial \alpha} = q\Phi(X) - C_\alpha(q, 0, \alpha). \quad (29)$$

The document supplier can only be detected if the chosen q and α are both strictly positive. This requires

$$rX + (1 + \alpha) \Phi(X) - C_q(q, 0, \alpha) \geq 0, \quad (30)$$

$$q\Phi(X) - C_\alpha(q, 0, \alpha) \geq 0. \quad (31)$$

To aid discussion, assume initially that the marginal monitoring costs are identical with $C_q(q, 0, \alpha) = C_\alpha(q, 0, \alpha)$. If there is an interior solution for q ($0 < q < 1$) then (30) is satisfied with equality. Since $\Phi(X) > 0$ (31) cannot be an equality, and so $\alpha = 0$, *i.e.* the optimal solution involves no attempt to catch intermediaries. The optimal value for α can only be positive if $q = 1$. Hence, the government finds it optimal to monitor document suppliers only when fraud by the firms is detected with certainty. But if $q = 1$ the firms would not engage in fraud in the first place. When the assumption of identical marginal monitoring costs is relaxed positive values of both q and α are possible only when $C_\alpha(q, 0, \alpha) < C_q(q, 0, \alpha)$. Hence, unless the marginal cost of monitoring intermediaries is relatively low the government will concentrate resources on auditing firms.

Now consider the scenario when the intermediary is a corrupt official and can be monitored separately from the firm. This makes p an additional choice variable. The necessary condition is obtained by substituting (8) into (24) and differentiating with respect to p . The result implies that p will be positive only if

$$(1 + \alpha) \Phi(X) - C_p(q, p, \alpha) \geq 0. \quad (32)$$

Assume initially that the marginal costs of monitoring are identical for q and p , so

$$C_q(q, p, \alpha) = C_p(q, p, \alpha). \quad (33)$$

Under this assumption it follows from (30) and (32) that if there is an interior solution for q ($0 < q < 1$) then p would necessarily be zero. Once more, this is a case for which resources would be focussed on auditing the firm. The value of p can be positive only if $q = 1$ – a situation in which there would be no fraud. Interior solutions for both audit probabilities can arise when the marginal cost of monitoring the intermediary is lower than the marginal cost of monitoring the firm. In general, when the relative marginal cost of monitoring the document suppliers is sufficiently high it is optimal for the government to monitor only the firm.

The results of this section have demonstrated that the possibility of tax fraud should not distort either the tax policy of the government or the level of rebate given. The analysis of monitoring shows that there is a strong incentive

to focus resources on catching dishonest firms. Even if the marginal cost of increasing the probability of catching firms is equal to the marginal cost of increasing the probability of catching intermediaries the optimal policy is still to focus resources on catching firms rather than intermediaries.

6 Fine on Tax Evaded

This section briefly reviews the implications of assuming that the penalty is determined by the value of fraud. Consequently, the penalty is now given by $\Phi = \Phi(E)$. Our main observation is that this change does not significantly affect the previous conclusions.

The price for the false documents, V , changes only to the extent that $\Phi(X)$ is replaced by $\Phi(E)$ in (11). The sufficient condition that some fraud is undertaken is given by

$$1 - q_F - (q_I + q_F) \Phi'(0) > 0, \quad (34)$$

which requires that the probabilities of the fraud being detected and the marginal rate of increase of the fine at zero fraud must not be too great. When fraud is undertaken the first-order condition for the choice of E gives

$$\Phi'(E) = \frac{1 - q_F}{q_I + q_F} = \frac{1 - q - \alpha p}{(1 + \alpha)(p + q)}. \quad (35)$$

The details of the comparative static analysis are given in the Appendix. To summarize, we show that the effects of the probability parameters, q , p , and α have the same sign as in the previous case. However, there is an additional result that the value of the excess claim is independent of the tax and rebate rates. The production and export decisions of the firm remain unaffected by the possibility of fraud.

The result in (35) shows that E is constant as t and/or r change. The objective function of the government can be written as

$$E[TR_F] = TR_H - (1 - q_F)E + (q_I + q_F)\Phi(E) - C(q, p, \alpha). \quad (36)$$

Since E is constant this differs only by a constant from the objective in the problem

$$\max_{\{t, r\}} TR_H, \quad (37)$$

which is the optimization of tax revenue with no fraud and the correct rebate for export production. The optimization (37) must therefore have the same optimal solution as in the case of honesty. Hence the possibility of fraud does not affect the optimal tax and rebate rates. Note that even though fraud does not change the optimal r and t it does affect the level of revenue.

The effect of the auditing probabilities upon the revenues net of costs is given by (30), (31) and (32) but with $\Phi(X)$ replaced by $\Phi(E)$. The same arguments then apply regarding the relative monitoring of the firm and the intermediary.

7 Convexity of Punishment

The Chinese criminal law relating to tax fraud is characterized by the fact that the punishments increase more than proportionally with the level of fraud. In extreme cases even capital punishment is possible. Such a punishment structure can be interpreted as being a convex function of the amount of fraud, and it is interesting to consider whether convexity of punishment emerges endogenously as the revenue-maximizing choice of the government. Kolm (1973) provided an argument why extreme punishments may have a role, but recent work by Boadway and Sato (2000) on endogenous punishment has shown more reasonable outcomes are also possible.

We begin by describing the outcome if the penalty function is concave. Note that

$$\frac{\partial^2 E[\pi_F]}{\partial X^2} = -\theta(q_I + q_F)\Phi''(X), \quad (38)$$

when the fine is determined by the extent of fraud, and

$$\frac{\partial^2 E[\pi_F]}{\partial X^2} = -\theta r^2(q_I + q_F)\Phi''(rX), \quad (39)$$

when the fine is determined by the value of fraud. In both cases the optimal level of X must be a corner solution if the fine is concave ($\Phi'' < 0$). In other words, with a concave punishment if there is an incentive for the firm to cheat at all then the expected gain from cheating is unbounded, and the firm will choose the maximal possible level of fraud unless it is monitored with certainty. Hence, if the government aims to reduce the amount of fraud when full monitoring of the firm is prohibitively expensive, it will employ a convex punishment schedule.

To investigate the issue of convexity further we adopt specific functional forms for Φ and C . This allows the optimal choice of the shape of the punishment function and the probabilities of fraud detection to be analyzed explicitly. We choose to focus on the case when the penalty is determined by the amount of tax evaded, $\Phi(E)$. The results are similar in the case of $\Phi = \Phi(X)$.

The functional forms that we adopt are

$$C = \frac{1}{2}(c_p p^2 + c_q q^2 + c_\alpha \alpha^2), \quad c_p, c_q, c_\alpha > 0, \quad (40)$$

and

$$\Phi = \phi E^\gamma, \quad \gamma > 0. \quad (41)$$

The positive constant ϕ is a scale parameter. γ is a choice parameter for the government, and the punishment is convex if $\gamma \geq 1$. With the assumed penalty function the expected gain from tax fraud for the firm is

$$E[\pi_F] - \pi_H = \theta(1 - q_F)[E - \phi E^\gamma], \quad (42)$$

where

$$\phi \equiv \frac{q_I + q_F}{1 - q_F} \phi. \quad (43)$$

Consider the choice of the firm if the governments selects a concave punishment function with $0 \leq \gamma < 1$. In this case E grows faster than φE^γ , and the two functions cross at

$$E^* = \varphi (E^*)^\gamma = \varphi^{1/(1-\gamma)} \quad (44)$$

so that for $E \geq E^*$ the expected gain monotonically increases in E . Consequently, for any finite value of φ , the expected benefit is unbounded. The firm therefore has an incentive to cheat without limit. If the objective of the government is to control fraud it therefore has to choose a convex penalty with $\gamma \geq 1$.

The optimal value of γ is determined by the magnitude of φ . The optimization can be analyzed by first observing that if $\varphi \geq 1$, the expected gain for the firm from fraud is non-positive for any $E > 0$ when $\gamma \geq 1$. The firm will therefore choose $E = 0$. The government can achieve this outcome by choosing low values of p , q , and α , and, provided that it is free to choose the value of ϕ , set ϕ high enough to ensure $\varphi > 1$. In other words, the government completely prevents tax fraud by announcing a linear or convex punishment with a high scale factor and performs the minimal auditing needed to ensure compliance. This is a similar conclusion to the one obtained by Kolm (1973) in the context of personal income taxation.

The value of ϕ sets the scale of the punishment function whereas γ controls the rate of increase of the gradient. There are several good reasons why there may be a limit on the value of the scale parameter, ϕ . This scale parameter can be interpreted as the punishment for a unit of fraud ($E = 1$). The standard approach to the economics of crime argues that the punishment for any illegal act must be set in line with those for other crimes to provide the correct incentives for the efficient minimization of crime (Cooter and Ulen, 2000). Setting the punishment at a high level for a limited quantity of tax fraud can be inconsistent with these incentives. This provides an argument for the limitation of fines for small amounts of tax fraud as part of a general crime prevention policy. We now choose to work under the assumption that there is an exogenous upper bound $\bar{\phi}$ on this parameter that prevents the government simply eliminating fraud while setting low values of the audit probabilities.

The government chooses (γ, p, q, α) to maximize

$$\begin{aligned} E[TR_F] - TR_H = & -(1 - (q + \alpha p)) E(\gamma, p, q, \alpha) \\ & + (1 + \alpha)(p + q) \bar{\phi} [E(\gamma, p, q, \alpha)]^\gamma - \frac{1}{2} (c_p p^2 + c_q q^2 + c_\alpha \alpha^2). \end{aligned} \quad (45)$$

subject to the constraint $0 \leq p, q, \alpha \leq 1$, $\gamma \geq 1$ and

$$E(\gamma, p, q, \alpha) = 0 \quad \text{if} \quad \frac{(1 + \alpha)(p + q)}{1 - p - \alpha q} \bar{\phi} \geq 1 \quad (46)$$

$$E(\gamma, p, q, \alpha) = \left[\frac{1 - (q + \alpha p)}{\gamma \bar{\phi} (1 + \alpha)(p + q)} \right]^{1/(\gamma-1)} \quad \text{if} \quad \frac{(1 + \alpha)(p + q)}{1 - p - \alpha q} \bar{\phi} < 1 \quad (47)$$

The expression for E in (47) is obtained by combining (8), (43), and (44).

In the tables we illustrate the optimal policy when $c_p = c_q = c_\alpha = 3$ for various levels of $\bar{\phi}$. Table 1 represents the case of the intermediary belonging to the hidden economy and, hence, cannot be monitored independently ($p = 0$). In Table 2 the intermediary is a corrupt official so can be monitored independently from the firm (p is a choice variable). The optimal policy involves auditing both the firm and the intermediary, with the probability of auditing the intermediary being significantly lower (this, in particular, results in the marginal cost of monitoring the intermediary being lower than the marginal cost of monitoring the firm). When the intermediary is a corrupt official, it is less than half as likely to be detected as is the firm. When the intermediary belongs to the underground economy the difference is even larger, and the difference increases the lower is the ceiling on the punishment scale factor, $\bar{\phi}$. Lower $\bar{\phi}$ also results in the punishment function being more convex, and it has opposite effects on the probabilities of monitoring the firms and the intermediaries: q_F increases, and q_I (slightly) decreases as $\bar{\phi}$ becomes lower. Hence, being able to levy only a small penalty on low amounts of fraud results in a rapidly increasing punishment structure.

$\bar{\phi}$	q_F	q_I	γ
0.05	0.78	0.07	2.01
0.10	0.76	0.09	1.61
0.15	0.75	0.10	1.37
0.20	0.74	0.12	1.23
0.25	0.73	0.13	1.12

Table 1: Optimal Policy
(Hidden Economy)

$\bar{\phi}$	p	q_F	q_I	γ
0.05	0.16	0.66	0.26	2.88
0.10	0.18	0.64	0.28	2.30
0.15	0.20	0.63	0.30	1.93
0.20	0.21	0.62	0.31	1.69
0.25	0.22	0.61	0.33	1.49

Table 2: Optimal Policy
(Corrupt Official)

8 Conclusions

The paper has provided an analysis of VAT fraud in the context of the Chinese rebate for exported goods. The interesting features of this situation are the convex punishment schedules which are defined in the relevant tax law and the fact that an additional party has to be involved in the supply of the documentation required to support the fraud.

The analysis shows that there is a well-defined choice problem even for a firm which maximizes profit. This is due to the convexity of punishments in the level of fraud. It is therefore unnecessary to impose risk aversion upon the preferences of the firm. The decision process of the firm separates into two parts. The real decisions of the firm – on input use and the level of output – are unaffected by the possibility of engaging in fraud.

Two options for the punishment schedules were considered: based on the extent of the fraud and based on the value of fraud. In the first case the extent of fraud increases as the tax rate increases. In the second case the value of the fraud remains constant as the tax rate is changed. The interesting finding is that in both cases the optimal tax rate is unaffected by the existence of fraud.

Hence, there is no incentive for the government to reduce the tax rate to reduce fraud or of raising it to cover for lost revenue. Fraud therefore does not imply a deviation of the tax rate from its efficient value.

The choice between focusing audit policy upon the firm or upon the supplier of false documentation depends primarily upon the cost of monitoring. If the marginal costs of monitoring the firm and the document supplier are equal, the policy is most likely to involve audits focused upon the firm. The analysis showed that when the punishment function is concave the firm has an incentive to commit the maximum possible fraud unless it is monitored with certainty. Therefore, a government with limited resources to monitor compliance will employ a convex punishment in order to reduce fraud. This result provides support for the structure of Chinese criminal law and shows that auditing and convex punishments can be determined endogenously in this model of crime.

We must note that the convex punishment structure for tax fraud in China should not, perhaps, be viewed as a unique phenomenon. Shikalova (2004) analyzed payroll tax evasion by firms in Russia, under a slightly different set of assumptions. Model calibration to Russian data showed that the shape of the punishment function for the payroll tax evasion is convex in the amount of labour income in “black cash”.

A Comparative static analysis

In the Appendix we list the properties of the optimal evasion decision of the firm summarized in Sections 4 and 6.

A.1 Evasion decision when time determined by the extent of fraud

$$\begin{aligned}
\frac{dX}{dt} &= 0, \\
\frac{dX}{dr} &= \frac{1 - q_F}{(q_I + q_F) \Phi''(X)} > 0 \\
\frac{dX}{dq} &= -\frac{1 + (1 - \alpha)p}{(1 + \alpha)(p + q)^2 \Phi''(X)} r < 0, \\
\frac{dX}{dp} &= -\frac{1 - (1 - \alpha)q}{(1 + \alpha)(p + q)^2 \Phi''(X)} r < 0, \\
\frac{dX}{d\alpha} &= -\frac{1 - q + p}{(1 + \alpha)^2 (p + q) \Phi''(X)} r < 0
\end{aligned}$$

A.2 Evasion decision when fine determined by the evaded tax

$$\begin{aligned}
\frac{dX}{dt} &= \frac{1 - q_F}{(q_I + q_F) \rho \Phi''(E)} \left(1 - \frac{p_h}{p_e} \right) > 0, \\
\frac{dX}{dr} &= \frac{1 - q_F}{(q_I + q_F) \rho \Phi''(E)} > 0, \\
\frac{dE}{dt} &= \frac{dE}{dr} = 0 \\
\frac{dX}{dq} &= -\frac{1 + (1 - \alpha)p}{(1 + \alpha)(p + q)^2} \rho < 0, \\
\frac{dX}{dp} &= -\frac{1 - (1 - \alpha)q}{(1 + \alpha)(p + q)^2 \Phi''(E)} < 0, \\
\frac{dX}{d\alpha} &= -\frac{1 - q + p}{(1 + \alpha)^2 (p + q) \Phi''(E)} < 0
\end{aligned}$$

References

- Ahmad, E., Singh, R., and Lockwood, B. (2004) "Taxation reforms and changes in revenue assignments in China," *IMF Working Paper*, No. 04/125.
- Aizenman, J. and Jinjark, Y. (2008) "The collection efficiency of the Value Added Tax: Theory and international evidence," *The Journal of International Trade and Economic Development*, **17**, 391-410.
- Allingham, M.G. and Sandmo, A. (1972) "Income tax evasion: a theoretical analysis", *Journal of Public Economics*, **1**, 323 - 338.
- Arrow, K.J. (1971) *Essays in the Theory of Risk Bearing*, Chicago: Markham Publishing Co.
- Boadway, R. and Sato, M. (2000) "The optimality of punishing only the innocent: the case of tax evasion", *International Tax and Public Finance*, **7**, 641 - 664.
- Chao, C., Yu, E., and Yu, W. (2006) "China's import duty drawback and VAT rebate policies: A general equilibrium analysis," *China Economic Review*, **17**, 432-448.
- Cooter, R. and Ulen, T. (2000) *Law & Economics*, Reading: Addison Wesley Longman.
- Cremer, H. and Gahvari, F. (1993) "Tax evasion and optimal commodity taxation", *Journal of Public Economics*, **50**, 261 - 275.
- Crocker, K.J. and Slemrod, J. (2005) "Corporate tax evasion with agency costs", *Journal of Public Economics*, **89**, 1593 - 1610.

- Das-Gupta, A. and Gang, I. (2003) "Value added tax evasion: Auditing and transaction matching," in: John McLaren (Ed.), *Institutional Elements of Tax Design and Reform*, World Bank Publications, 25-48.
- de Mello, L. (2008), "Avoiding the value added tax: Theory and cross-country evidence", *OECD Economics Department Working Papers*, No. 604, OECD Publishing. doi:10.1787/242167601452, forthcoming in the *Public Finance Review*.
- Fedeli, S. and Forte, F. "Joint income-tax and VAT-chain evasion," *European Journal of Political Economy*, **15**, 391-415.
- Fisman, R. and Wei, S.-J. (2001) "Tax rates and tax evasion: Evidence from 'missing import' in China," *NBER Working Paper*, No. 8551.
- Jia, S. (2002) *Zhongguo Shuishou Lliushi Wenti Yanjiu (Analysis of Tax Evasion Problem in China)*, Zhogguo Caizheng Jingji (Chinese Finance and Economics Press).
- Keen, M. and Mintz, J.(2004) "The optimal threshold for a value-added tax," *Journal of Public Economics*, **88**, 559 - 576.
- Keen, M. and Smith, S. (2006) "VAT fraud and evasion: What do we know and what can be done?" *National Tax Journal*, **109**, 861-887.
- Keen, M. (2008) "VAT, tariffs, and withholding: Border taxes and informality in developing countries," *Journal of Public Economics*, **92**, 1892–1906.
- Kolm, S.C. (1973) "A note on optimum tax evasion", *Journal of Public Economics*, **2**, 265 - 270.
- Lin, S. (2008) "China's value-added tax reform, capital accumulation, and welfare implications," *China Economic Review*, **19**, 197-214.
- Marrelli, M. (1984) "On indirect tax evasion", *Journal of Public Economics*, **25**, 181 - 196.
- Marrelli, M. and Martina, R. (1988) "Tax evasion and strategic behaviour of the firms", *Journal of Public Economics*, **37**, 55 - 69.
- Myles, G.D. (1995) *Public Economics*, Cambridge: Cambridge University Press.
- Reinganum, J. and Wilde, L. (1986) "Equilibrium verification and reporting policies in a model of tax compliance", *International Economic Review*, **27**, 739 - 760.
- SAT (2003, 2005) <http://www.gov.cn>.
- Schneider, F. and Enste, D.H. (2000) "Shadow economies: size, causes, and consequences", *Journal of Economic Literature*, **38**, 77 - 114.

- Shikalova, M. (2004) “Effect of pension reform on tax evasion: Russian case”, *mimeo*, Department of Economics University of Oslo, 20 pp.
- Slemrod, J. and Yitzhaki, S. (2000), “Tax avoidance, evasion, and administration”, *NBER Working Paper*, No. 7473.
- Stöwhase, S. (2005), “Asymmetric capital tax competition with profit shifting”, *Journal of Economics*, **85**, 175-196.
- Stöwhase, S. and Traxler, C. (2005) “Tax evasion and auditing in a federal economy”, *International Tax and Public Finance*, **12**, 515 - 531.
- Virmani, A. (1989) “Indirect tax evasion and production efficiency”, *Journal of Public Economics*, **39**, 223 - 237.
- Whalley, J. and Wang, L. (2007) “Evaluating the impure Chinese VAT relative to a pure form in a simple monetary trade model with an endogenous trade surplus”, *NBER Working Paper*, No. 13581.
- Yamada, M. (1990) “An analysis of optimal taxation with tax evasion”, *Public Finance*, **45**, 470 - 490.